

T-15 複数行列の行列分解に関する一考察

幸島 匡宏, 江崎 健司, 高屋 典子, 澤田 宏(NTT)

- ノイズが含まれた複数のデータから, データ間の関係を見出す問題
- **ノイズ**を陽に抽出したうえで**クラスタ(トピック)**形成を行う複数行列の非負値行列分解手法を提案

$$\begin{array}{ccc} I \times J & & I \times R^{(1)} \quad R^{(1)} \times J \\ \boxed{X_1} & \approx & \boxed{A_1} \quad \boxed{B^T} \\ & & \text{green bar} \quad \text{grey box} \end{array} \quad \begin{array}{ccc} I \times K & & I \times R^{(2)} \quad R^{(2)} \times K \\ \boxed{X_2} & \approx & \boxed{A_1 A_2} \quad \boxed{C^T} \\ & & \text{green bar} \quad \text{red bar} \quad \text{grey box} \end{array}$$

- 20Newsgroupデータ★への適用から, ラベル推定精度の向上と **stopwords** を分離したトピック形成が可能となることを確認

★ 20Newsから作成するデータ

- X_1 : 文書-ラベル行列
- X_2 : 文書-単語行列

topic1	apple drive mac meg cable
topic2	window motif server mit code
topic3	team ca hockey play game
topic4	key faq security public sci
topic5	key encryption government clipper keys
topic*	the to of and in